SVC-3003941

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

APPLICATION FOR LETTERS PATENT

INVENTOR:

Dey et al.

TITLE:

Tool for Automatically Mapping Multimedia Annotations to Ontologies

BACKGROUND OF THE INVENTION

Field of Invention

The present invention relates generally to the field of multimedia (video, audio, graphics, etc.) presentations authoring. More specifically, the present invention is related to intelligently integrating multimedia content and other contextually related content via an associative mapping system.

Discussion of Prior Art

Definitions have been included to help with a general understanding of associative mapping terminology and are not meant to limit their interpretation or use thereof. Other definitions or equivalents may be substituted without departing from the scope of the present invention.

Annotation: A comment attached to a particular section of a document. Many computer applications enable a user to enter annotations on text documents, spreadsheets, presentations, images, and other objects. It should be noted that the terms "annotation" and "keyword" equivalent and are therefore used interchangeable throughout the specification.

Ontology: The hierarchical structuring of knowledge about objects by sub-categorizing based on their relevant qualities.

The following references describe prior art in the field of associate mappers. The prior art mentioned below describe associative mapping in general, but none provide the benefits of the

present invention's method and system for automatically mapping multimedia document annotations (or keywords) to ontologies.

US Patent 5,056,021 to Ausborn provides for a method and apparatus for abstracting concepts from natural language, wherein each word is analyzed for its semantic content by mapping into its category of meanings within each of four levels of abstraction. Each word is mapped into the various levels of abstraction, forming a file of category of meanings for each of the words. This is a manual process done by knowledge engineers prior to using this file for abstracting meanings from natural language words.

US Patent 6,061,675 to Wical provides for a method and apparatus for classifying terminology utilizing a knowledge catalog, wherein the static ontologies store all senses for each word and concept giving a broad coverage of concepts that define knowledge. A knowledge catalog processor accesses the knowledge catalog to classify input terminology based on the knowledge concepts in the knowledge catalog.

These prior art systems are not very suitable for automatically learning to relate loosely defined or unstructured contextual information (such as annotations or keywords or captions or transcripts) of a multimedia document sequence to formally or semi-formally represented ontologies related to sequences of multimedia documents. The following are some of the main problems associated with conventional associative mappers:

- The process of building the catalog or indices is not automatic and needs elaborate human engineering to attach the words to concepts or nodes in the ontology (or taxonomy, interchangeably used from hereon).

- In the domain of mapping multimedia document annotations, prior engineering of words by attaching them to concepts in the ontology is not feasible due to the drifting nature of the relevance of words to concepts in the ontology.

- Conventional associative mappers do not deal with groups of words (as in annotations) that occur together (and not a full natural language sentence), and hence lead to issues like topic cross talk (described in detail later). Annotations in multimedia documents usually tend to be about more than one topic. This leads to problems in learning from data derived from past annotation mappings.

- Conventional associative mappers rely on natural language processing systems that require more processing.

Associative mappers described in prior art systems fail to provide for a multimedia document authoring environment that helps rapidly create a document that integrates multimedia content with other content that is relevant to a segment of the multimedia document. Furthermore, prior art

systems fail to describe an information retrieval mechanism that intelligently combines and renders multimedia content with other contextual content via a server on a network.

In these respects, the tool for mapping multimedia document annotations to ontologies according to the present invention substantially departs from the conventional concepts and designs of the prior art. Thus, it provides an apparatus primarily developed for the purpose of learning to map annotations or captioning of multimedia documents to nodes or concepts in formally or semi-formally represented ontologies covering a broad range of possible multimedia documents.

Whatever the precise merits, features and advantages of the above cited references, none of them achieve or fulfill the purposes of the present invention.

SUMMARY OF THE INVENTION

A tool is introduced for automatically mapping multimedia annotations to ontologies wherein the same is utilized for learning to relate annotations or captioning of a multimedia document to nodes or concepts in formally or semi-formally represented ontologies covering a broad range of possible multimedia documents. Therefore, the associative mapper of the present invention provides for a multimedia document authoring environment that helps rapidly create a document that integrates multimedia content with other content that is relevant to the multimedia segment. Furthermore, the associative mapper of the present invention is used in conjunction with a server in a network to render an integrated presentation comprising multimedia document and other

contextually related content.

The key components of the system of the present invention include:

1. Learning data preparation component that involves techniques for deriving data from past mappings of annotations (or keywords) to nodes in a taxonomy or an ontology. Learning represents the ability of a device to improve its performance based on the past performance data;

2. Intelligent inverted indices component maintaining statistics, and

3. A retriever that exploits these statistics to rank the relevance of the nodes in a taxonomy for a given set of new annotations.

The above-mentioned learning data preparation component, intelligent inverted index component or IIndex (for maintaining certain special statistics), and a retriever (that exploits the statistics maintained by IIndex to rank the relevance of the nodes in a taxonomy for given a set of new annotations) form the main components of this invention. Thus, the present invention provides for a technology for automatic and dynamic mapping of multimedia documents to ontologies via the three components described above.

Thus, the more important features of the present invention have been outlined, rather broadly, in order that the detailed description thereof may be better understood and that the present contribution to the art may be better appreciated. There are additional features of the invention that will be described hereinafter.

Other advantages of the present invention will become obvious to the reader and it is intended that these advantages are within the scope of the present invention.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1a illustrates an overview of the learning data component associated with the system of the present invention.

Figure 1b illustrates an example of mapped nodes in a taxonomy.

Figure 2 illustrates an overview of the method associated with the system in Figure 1.

Figure 3 illustrates the method associated with learning data preparation.

Figure 4 illustrates a statistical calculation maintained by the IIndex of the system of the present invention.

Figure 5 illustrates a graph of a second component associated with the weighting factor *wt_cf*.

Figure 6 illustrates a statistical calculation maintained by the retriever component of the system of the present invention.

Figure 7 illustrates the method associated with the interactive multimedia document authoring environment.

Figure 8 illustrates ways of obtaining various multimedia document annotations.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

While this invention is illustrated and described in a preferred embodiment, the invention may be produced in many different configurations, forms and materials. There is depicted in the drawings, and will herein be described in detail, a preferred embodiment of the invention, with the understanding that the present disclosure is to be considered as an exemplification of the principles of the invention and the associated functional specifications for its construction and is not intended to limit the invention to the embodiment illustrated. Those skilled in the art will envision many other possible variations within the scope of the present invention. Furthermore, it is to be understood that the phraseology and terminology employed herein are for the purpose of the description and should not be regarded as limiting.

Figure 1a illustrates an overview of components associated with the system of the present invention. A learning data preparation component looks at the annotations (e.g., multimedia annotations **102**) and their past mappings into the nodes in the taxonomy and prepares the learning instances, one per node in the taxonomy. Figure 1b illustrates an example of mapped nodes in a taxonomy. In this example, the "Boston" node is linked to three nodes: "Boston Red Sox", New England Patriots", and "Boston Globe". But, the "Boston Red Sox" node is also linked to the "Baseball Teams" node (and so is the "New York Yankees" node), and similarly the "Boston Globe" node is also linked to the "Newspapers" node. Furthermore, the "Boston" node is also linked to the

"Major US Cities" node. Lastly, the "Pedro Martinez" node is linked to the "Boston Red Sox" node.

Returning to the discussion in Figure 1a, the prepared learning instances are tokenized (via tokenizer **104**), stemmed **106**, stop words are removed **108**, and passed on to the IIndex **110**. This component generates *tf*, *idf* and *cf* statistics for the learning instances (from learning data prepared from annotations **112**) and creates an inverted index that is a data structure that maps words to nodes to which those words are associated.

Thus, the learning data preparation occurs prior to the search process. During the search process, the retriever looks at new annotations and uses the inverted index to retrieve and rank most relevant nodes for these annotations. The ranking process uses equations 1, 2, 3, and 4 (discussed below) to calculate the weights and rank the nodes (thereby forming ranked topics **114**) in the order of their relevance.

Figure 2 illustrates an overview of the method **200** associated with the system in Figure 1, wherein the learning data preparation component looks at the annotations and their past mappings, to the nodes in the taxonomy and prepares the learning instances **202**, one per node in the taxonomy. IIndex treats these learning instances as a bag of words to be indexed and generates *tf*, *idf* and *cf* statistics for them and creates an inverted index **204**. During the search process, the retriever looks at new annotations and uses the inverted index to retrieve and rank most relevant concepts from the ontology **206**.

A detailed description of the above described learning system, intelligent inverted index, and retriever mechanisms are provided below:

Learning data preparation:

Learning represents the ability of a system or device to improve its performance based on past performance data. A learning system has to be endowed with the capability to look at the past performance data and derive abstract patterns of regularities that are generalized to novel situations. Learning data preparation, as illustrated in Figure 3, involves looking at the data derived from past mappings of annotations and captions to the ontology **300** and fusing all annotations that are mapped into the same node in the ontology into a learning instance for that node **302**. The fused annotations make words relevant to the node standout more than in individual annotations. Such a fusing also solves the problems of "short documents" that lead to poor results when using classical information retrieval techniques. Fusing annotations also lead to lesser sensitivity to errors in mappings. One of the most significant gains from fusing annotations mapped to a node for forming a learning instance vector is the mitigation of the topic cross talk problem. Supposing the annotations associated with topics "basketball" and "shoes" are detailed and long, where as those that are associated with "basketball" and "injury" are sparse and short. Then, a query associated with "basketball" and "injury" is likely to lead to the retrieval of the nodes related to "shoes" because of high term-frequencies for terms related to "basketball" and "shoes" in these annotations and low term-frequencies for terms related to "basketball" and "injury" annotations. This phenomenon is defined

as "topic cross talk". Each annotation is associated with more than one topic. Hence, words related to more than one particular topic occur in an annotation and get associated with that topic. Later, a discussion of the details of the mitigation of topic cross talk is provided. It relies on a statistical mechanism called "contribution frequency" that relies on the fused annotations.

Intelligent Inverted Index for maintaining certain special statistics:

IIndex starts with standard information retrieval (IR) technology (for building inverted indices for unstructured information) and incorporates a number of enhancements to make it effective for the task of relating annotations and captioning to nodes in a taxonomy. Standard IR systems rely on building an inverted index that is a data structure that maps words to documents in which those words occur. In addition, the inverted index also maintains certain statistics like term frequency (*tf*) and inverse document frequency (*idf*) for the words and their corresponding documents. Term frequency tf_{ij} is the number of times a particular word *i* occurs in a document *j*. Document frequency df_i represents the number of documents in the entire document database in which the word *i* occurs at least once. As shown in Figure 3, the system of the present invention relies on these statistics and augments them with a novel statistic called "contribution frequency", denoted by *cf*, that is particularly suited to avoid topic cross talk in learning instances derived from fused annotations. For each word in a fused learning instance, its *cf* is just the number of annotations (that comprise the instance) in which the word appears. The statistic *tc* is the total number of annotations that comprise that learning instance.

Furthermore, Figure 4 illustrates a statistical calculation maintained by the IIndex of the system of the present invention. Standard statistical calculations like inverse document frequency (*idf*), term frequency (*tf*), and document frequency (*df*) are identified in step **400**. Next, two of the above-described statistics: contribution frequency (*cf*) and total number of annotations (*tc*) are identified in step **402**. In step **404**, a weighting factor (*wt_cf*) with regard to the contribution frequency (*cf*) is calculated.

The weighting factor *wt_cf*, is calculated based on:

$$wt\_cf = \underbrace{(0.5+\frac{cf}{tc})}_{\text{Component 1}} \quad \underbrace{(1.0-\frac{0.5}{1+0.05tc^2})}_{\text{Component 2}}$$

The *wt_cf* measure consists of two components. The first component takes care of the fact that the higher the *cf* with respect to *tc*, the higher the *wt_cf*. Thus, the higher the contribution frequency of a word to a particular concept, then the higher its weight in determining the relevance of the concept. The addition of constant 0.5 makes *wt_cf* less sensitive to this ratio. The second component has a functional form as in Figure 5. This component takes on the role of assigning fewer weights to the evidence derived from the *cf/tf* ratio when the number of abstracts comprising a learning instance is small. In other words, occurring in 2 abstracts out of 5 total abstracts in a topic document is not the same as occurring in 20 abstracts out of 50. The evidence in the latter case is stronger. However, once the total abstracts is more than about 30 (this parameter was experimentally determined to be optimal for the domain of multimedia annotation mapping), the second component levels off at 1.0.

Retriever mechanism to exploit the special statistics maintained by IIndex:

The retriever exploits the special statistic maintained by IIndex to rank the relevance of the nodes in a taxonomy for given set of new annotations. The retrieval mechanism uses the same measures as the intelligent indexing mechanisms that IIndex uses. It relies on *tf*, *idf* and *cf* and uses Equations 1, 2, 3, and 4 (given below) to rank the retrieved nodes in their order of relevance to a new annotation. Figure 6 illustrates the statistical calculations performed by the retrieval mechanism. Contribution of the term frequency to the weight of a query term ($Normalized\_tf_{ij}$) is calculated in step **602** (Equation 1). In step **604**, an inverse document frequency (*idf*) is calculated, wherein the *idf* is normalized with respect to the number of documents (Equation 2). Lastly, a calculation is performed, as in step **606**, to identify the weight contributed to a particular category in the ontology by the occurrence of word i in learning vector j (Equation 4).

Equation 1:

$$Normalized\_tf_{ij} = 0.4 + 0.6 \times \frac{\log(tf_{ij} + 0.5)}{\log(\max\_tf_j + 1)}$$

Equation 2:

$$idf_i = \frac{\log(\frac{N}{df_i})}{\log(N)},$$

where "N" is the total number of documents.

Equation 3:

$$wt\_cf = (0.5 + \frac{cf}{tc}) \quad (1.0 - \frac{0.5}{1 + 0.05tc^2})$$

Equation 4:

$$wt_{ij} = (0.4 + 0.6 \times Normalized\_tf_{ij} xidf_j) \times wt\_cf$$

As stated earlier, term frequency "tf_{ij}" is the number of times a particular word *i* occurs in a document *j*. "max_tf_j" is the maximum term frequency of all the terms in document *j*. Document frequency df_i represents the number of documents in the entire document database in which the word *i* occurs at least once. The statistic, *cf*, is the number of annotations (that comprise the instance) in which the word appears. Furthermore, the statistic, *tc*, is the total number of annotations that comprise that learning instance. The statistic, *wt_cf*, is the weighting factor due to the contribution frequency. "wt_{ij}" is the weight contributed by the occurrence of word *i* in document *j*.

Equation 1 defines the contribution of the term frequency to the weight of a query term. The fraction log (tf_{ij} + 0.5)/ log (max_tf_j + 1) defines normalized term frequency adjusted for the possibility of tf_{ij} being zero. The addition of small positive quantities to tf_{ij} and max_tf_j avoids applying log to a zero (this is undefined). The multiplicative constants 0.4 and the additive constant 0.6 reduce the sensitivity of $normalized\_tf_{ij}$ to the fraction $log(tf_{ij} + 0.5)/ log(max\_tf_j + 1)$. Equation

2 defines the inverse document frequency normalized by the total number of documents N. Equation 3 has been described previously with respect to Figure 5. Equation 4 takes the combined effects of normalized term frequency, inverse document frequency, and contribution frequency to arrive at the weight contributed to a particular category in the ontology by the occurrence of word *i* in learning vector *j*.

EXAMPLE IMPLEMENTATIONS

In one embodiment, the above-mentioned tool is part of a larger system that allows delivery of multimedia content integrated with other contextual content. This integrated experience is accessed via several devices, such as an interactive television, a computer, a telephone, a fax machine, or a handheld device, connected to the Internet, a cable system or a wireless network. Contextually related content is of several types: (i) text documents such as product bulletins, manuals, data sheets, press releases, news stories, biographies, analyst documents, (ii) message boards, chat rooms, (iii) product descriptions with instant purchase abilities (e-commerce), (iv) other multimedia documents consisting of audio, video, images and graphics in various formats, etc.

The system is unique in that it largely automates the end-to-end process of linking contextual content to multimedia presentations. Current systems allow a content producer to handcraft such an experience, leading to high resource requirements and lower productivity. We describe two major components of the system below:

A. Interactive Multimedia Authoring Environment:

The multimedia authoring environment enables a broadband producer to rapidly create a document that integrates multimedia content with other content that is relevant to the multimedia segment. Other relevant content resides on the Internet or within the intranet environment that the producer is in.

Currently, the producer would have to manually "attach" or "link" such content with the multimedia content. Figure 7 illustrates the method (**700**) associated with the interactive multimedia authoring environment wherein using the automatic mapping tool, the producer annotates the multimedia segment only **712**. Then the multimedia segment is automatically mapped to the appropriate node in the ontology **714**. Other related content that are mapped to the same node in the ontology are then to be integrated along with the multimedia segment **716**.

Producers have two options: They either (a) go through the related content, and pre-certify what is to be displayed to the viewer, or (b) allow dynamic content linking (described below).

Figure 8 illustrates some of the many ways to obtain annotations of the multimedia document **800**: (a) using existing closed captioning or a subset of it **802**, (b) using textual descriptions that accompany the multimedia document **804**, (c) by employing speech-to-text techniques **806**, and (d) by manually entering words that describe important aspects of a segment **808**.

B. Interactive Multimedia Delivery Server:

The Interactive Multimedia Delivery Server is responsible for presenting an integrated presentation consisting of multimedia and other contextually related content.

The unique architecture of this Interactive Multimedia Document Delivery Server is that the contextual information is not sent to user before it is requested (by the user). Whenever contextual information is needed by the end-user, the time within the multimedia document is used to determine the context within the presentation. Using this information, the server retrieves contextual information using searching it's own ontology and databases using Information Retrieval techniques, as well as sending queries to other databases and web sites. This dynamic content linking allows for information to be up-to-date as well as eliminate expired information.

Furthermore, the present invention includes a computer program code based product, which is a storage medium having program code stored therein, which can be used to instruct a computer to perform any of the methods associated with the present invention. The computer storage medium includes any of, but not limited to, the following: CD-ROM, DVD, magnetic tape, optical disc, hard drive, floppy disk, ferroelectric memory, flash memory, ferromagnetic memory, optical storage, charge coupled devices, magnetic or optical cards, smart cards, EEPROM, EPROM, RAM, ROM, DRAM, SRAM, SDRAM or any other appropriate static or dynamic memory, or data storage devices.

Implemented in computer program code based products are software modules for: receiving a request for searching and extracting one or more annotations related to said multimedia documents from an ontology; identifying nodes in the ontology that are relevant to the multimedia documents, wherein the nodes further comprises fused learning instances formed by fusing annotations based upon using statistics including term frequency, inverse document frequency and contribution frequency; and extracting information from said identified relevant nodes and dynamically linking said extracted information with said multimedia documents.

CONCLUSION

A system and method has been shown in the above embodiments for the effective implementation of a tool for automatically mapping multimedia annotations to ontologies. While various preferred embodiments have been shown and described, it will be understood that there is no intent to limit the invention by such disclosure, but rather, it is intended to cover all modifications and alternate constructions falling within the spirit and scope of the invention, as defined in the appended claims. For example, the present invention should not be limited by software/program, computing environment, or specific computing hardware.

The above enhancements for a method and a system for automatically mapping annotations of multimedia documents to ontologies and its described functional elements are implemented in various computing environments. For example, the present invention may be implemented on a conventional IBM PC or equivalent, multi-nodal system (e.g. LAN) or networking system (e.g. Internet, WWW, wireless web). All programming and data related thereto are stored in computer memory, static or dynamic, and may be retrieved by the user in any of: conventional computer storage, display (i.e. CRT) and/or hardcopy (i.e. printed) formats. The programming of the present invention may be implemented by one of skill in the art of statistical and network programming.